Steroids 74 (2009) 498–503

Contents lists available at ScienceDirect

Steroids

journal homepage: www.elsevier.com/locate/steroids

Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods☆

Hubert W. Vesper[a,*], Shalender Bhasin[b], Christina Wang[c], Susan S. Tai[d], Larry A. Dodge[e], Ravinder J. Singh[e], Judie Nelson[f], Susan Ohorodnik[g], Nigel J. Clarke[h], Wael A. Salameh[h], C. Richard Parker Jr.[i], Raj Razdan[a], Elizabeth A. Monsell[a], Gary L. Myers[a]

[a] Centers for Disease Control and Prevention, 4770 Buford Highway, NE F25, Atlanta, GA 30341-3724, United States
[b] Boston University, School of Medicine, 670 Albany Street, Boston, MA 02118, United States
[c] Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, UCLA School of Medicine, 1000 W. Carson Street, Torrance, CA 90509, United States
[d] National Institute of Standards and Technology, Chemical Science and Technology Laboratory, Analytical Chemistry Division, Gaithersburg, MD 20899, United States
[e] Mayo Foundation, 730 Hilton Building, Rochester, MN 55905, United States
[f] Newborn Screening/Biochemical Genetics Labs, Children's & Women's Health Centre of BC, 4480 Oak Street, Vancouver, BC V6H-3V4, United States
[g] Taylor Technology, Inc., 301D College Road East, Princeton, NJ 08540, United States
[h] QuestDiagnostics Nichols Institute, 33608 Ortega Highway, San Juan Capistrano, CA 92675, United States
[i] University of Alabama at Birmingham, Department of Obstetrics and Gynecology, 618 South 20th Street, Birmingham, AL 35294, United States

ARTICLE INFO

Article history:
Received 18 November 2008
Received in revised form 16 January 2009
Accepted 20 January 2009
Available online 30 January 2009

Keywords:
Testosterone
HPLC–MS/MS
Method comparison
Analytical variability

ABSTRACT

Background: Though mass spectrometry (MS) assays are increasingly used for routine clinical measurements of serum total testosterone (TT), information about the variability of results is limited. This study assessed the variability of TT measurement results from routine MS assays.
Methods: Twenty serum samples (12 females, 8 males) were analyzed on 2 days by seven high performance liquid chromatography (HPLC), and one gas chromatography (GC)–tandem mass spectrometry (HPLC–MS/MS, GC–MS/MS) assays. Two samples (male and female) were provided in five replicates to assess the within-run variability. Results were compared against those obtained at National Institute of Standards and Technology (NIST). The within- and between-laboratory variability was assessed for each sample. Comparisons to the NIST results were performed using bias plot and Deming regression analysis.
Results: The overall coefficient of variation of the results obtained with MS assays was <15%CV at >1.53 nmol/L and <34%CV at 0.3 nmol/L. The between-assay variability was the major contributor to the overall variability. The assay precision was the highest (<3%CV) with assays using liquid–liquid extraction for sample preparation or GC–MS/MS. The mean percent difference to the reference assay was 11%. The slopes of Deming regression analysis of the MS assays were between 0.903 and 1.138 (correlation coefficient: >0.996). TT concentrations for one assay were above the measurement range.
Conclusions: The variability of TT measurement results among MS assays is substantially smaller than that reported for immunoassays. The type of sample preparation may affect assay precision. Standardizing assays can further reduce the variability of measurement results.

Published by Elsevier Inc.

1. Introduction

Testosterone affects major biological functions. The type and magnitude of the effects are highly dependent on the concentration present in the target organs. Measuring circulating testosterone concentrations either as total or in non-protein-bound (free form) is used as a surrogate for target organ concentrations when investigating androgen status in children and adults of both sexes and for monitoring testosterone therapy as described in many position statements, recommendations and clinical guidelines [1–6].

Testosterone is commonly measured by direct immunoassays on automated, multipurpose clinical analyzers. The performance of immunoassays was examined in several studies [7–13]. These studies have revealed problems related to accuracy, precision, specificity and sensitivity, especially at lower concentrations commonly observed in women or children. Several expert panels expressed concerns about the reliability of platform-based immunoassay, especially at these low testosterone concentrations. Similarly, inaccuracies in the direct assays for free testosterone have led expert

☆ *Disclaimer*: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author. Tel.: +1 770 488 4191; fax: +1 404 683 2467.
E-mail address: HVesper@cdc.gov (H.W. Vesper).

0039-128X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.steroids.2009.01.004

panels to recommend against the use of these direct analog assays for free testosterone [6,7,13–23]. This situation profoundly hampers progress in research conducted on androgen-related disorders and limits the translation of existing research findings into clinical practice that would help improve patient care and disease prevention. Researchers and professional societies such as the Endocrine Society have suggested the need to standardize testosterone measurements to overcome these problems.

Mass spectrometry (MS)-based methods have specificity, sensitivity, and accuracy that seem to overcome the limitations observed with immunoassays and have therefore been recommended, especially for assessing testosterone concentrations in women and children [14,24,25]. Initial MS assays used gas chromatography coupled with mass spectrometry and were developed as reference methods [26–30]. However, the specimen requirements and sample throughput of these reference methods are not suited for routine clinical measurements. More recently, methods using high-performance liquid chromatography coupled with tandem mass spectrometry (HPLC–MS/MS) have been described and can be automated to provide high sample throughput in routine clinical measurements [31–37]. Current MS assays are developed, optimized and validated in-house using different procedures, instrumentation, reagents and calibrators. This can lead to differences in individual assay performances and in limitations when comparing results. Only a few studies assessed the accuracy of pairs of MS-based methods and found good agreement between methods [32–38]. The results from the most recent comparison study suggested that agreement and performance can further be improved through standardization [38].

The Centers for Disease control and Prevention (CDC) is currently working with the Endocrine Society, the American Association for Clinical Chemists, the American Association of Clinical Endocrinologists, the College of American Pathologists, and other organizations and institutions to standardize and improve steroid hormone measurements. As part of this effort, we investigated the variability of results obtained with different MS-based assays that are currently used for research and routine clinical testosterone measurements.

2. Experimental

Eight commercial and research laboratories participated in this study. Seven laboratories used high performance liquid chromatography tandem mass spectrometry (HPLC–MS/MS) assays and one gas chromatography tandem mass spectrometry (GC–MS/MS) assay. The basic characteristics of the assays are summarized in Table 1. The measurement results were compared against those obtained using the HPLC–MS/MS reference method performed by the National Institute of Standards and Technology (NIST). This method employs an extensive sample preparation consisting of a liquid–liquid and solid phase extraction, has a recovery of 100% and a within-run and between run imprecision of <1%CV [30]. This method is acknowledged by the Joint Committee for Traceability in Laboratory Medicine (JCTLM) as reference method. All results were converted to nmol/L and ng/dL units. The reference laboratory reported results in ng/g serum this unit was converted using an average density of serum of 1.02 g/mL.

Two sets of 30 fresh-frozen, blinded serum samples (1.2 mL each) were sent on dry ice to the laboratories. The samples were derived from 8 men (age range: 36–67 years) and 12 women (age range: 22–74 years). The participants were asked to analyze each set for total testosterone (TT) on different days with each run having its own calibration.

In each set of samples, two (TT concentration as reported by the NIST laboratory: 10.30 nmol/L [297 ng/dL] and 0.29 nmol/L [8.47 ng/dL], respectively) were provided in five replicates to assess the within-run variability. The average within-run variability

Table 1
Characteristics of the MS methods.

	Assay 1[a]	Assay 2	Assay 3	Assay 4	Assay 5	Assay 6	Assay 7	Assay 8
Instrument	HPLC–MS/MS	HPLC–MS/MS	HPLC–MS/MS	HPLC–MS/MS	HPLC–MS/MS	HPLC–MS/MS	HPLC–MS/MS	GC–MS/MS
System	Applied	Thermo Electron	Applied	Applied	Applied	Thermo Electron	Micromass	Thermo Electron
Manufacturer	Biosystems	TSQ Quantum	Biosystems	Biosystems	Biosystems	TSQ Quantum	Quattro Micro	TSQ 700 or 7000
Make	API 4000	Ultra	API 4000	API 4000	API-5000	Ultra		Varian model 3400 GC
Ionization	ESI	APCI	APCI	APCI	ESI	ESI	ESI	EI
Analyte transitions used (m/z)	289/97	Ni	289.4/97.3	289/109	289.2/109.0	289/97.1	289.10/97.10	535.1/505.1
Confirmation ions used (m/z)	No	Yes	289.4/108.9	Ni	Ni	289/109.1	No	No
Calibration range (nmol/L) (ng/dL)	0.277–30.06 (8.0–867)	0.035–69.35 (1–2000)	0.243–69.35 (7–2000)	0.087–173.50 (2.5–5000)	0.035–69.35 (1–2000)	0.06–969.35 (2–2000)	0.1–100 (14.41–2882)	0.087–11.10 (2.5–320)
Units reported	ng/g	ng/dL	ng/dL	ng/dL	ng/dL	ng/dL	nmol/L	pg/mL
Sample preparation	SPE and liquid–liquid extraction	On-line HTLC extraction	Protein precipitation and HTLC	Non-polar extraction, 2-column HPLC	Liquid–liquid extraction	Protein precipitation and HTLC	Liquid–liquid extraction	SPE, derivatization

Ni: no information provided.
[a] Reference method.

observed in both samples on 2 different days (n = 10 laboratory/per laboratory), expressed as %CV, was calculated using the pooled estimates of variance. The variability of individual results across assays (overall sample variability) reported for the same sample was calculated and expressed as %CV.

To assess the recovery of the assays, three samples were prepared by mixing a pool from a woman with the pool from a man in different ratios (75/25, 50/50, 25/75, v/v). The pools and mixed samples were added to each set. The recovery was calculated as the percent ratio of the measured and expected value. The difference of each assay to the NIST assay was determined by using the mean values obtained for each sample and by performing percent bias plot and weighted Deming regression analysis. The absolute value of the percent difference from each method was used to calculate the average percent difference of the methods to the reference method.

3. Results

The specific procedures used for measuring total testosterone in the serum samples were different for all assays (Table 1). Two laboratories reported using liquid–liquid extraction for sample preparation while all other laboratories used some form of solid phase extraction (off-line or on-line). Two laboratories performed protein precipitation prior to solid phase extraction. Five different sources of calibrators (Sigma, Steraloids, U.S. Pharmacopeia, Cambridge Isotope Laboratories, Inc., National Metrology Institute of Australia) and four different sources of internal standards (Sigma, CDN, Steraloids, Cambridge Isotope Laboratories, Inc.) were reported. All internal standards had deuterium as stable isotope label with the number of deuterated atoms ranging from two to five. Three laboratories reported using confirmation ions.

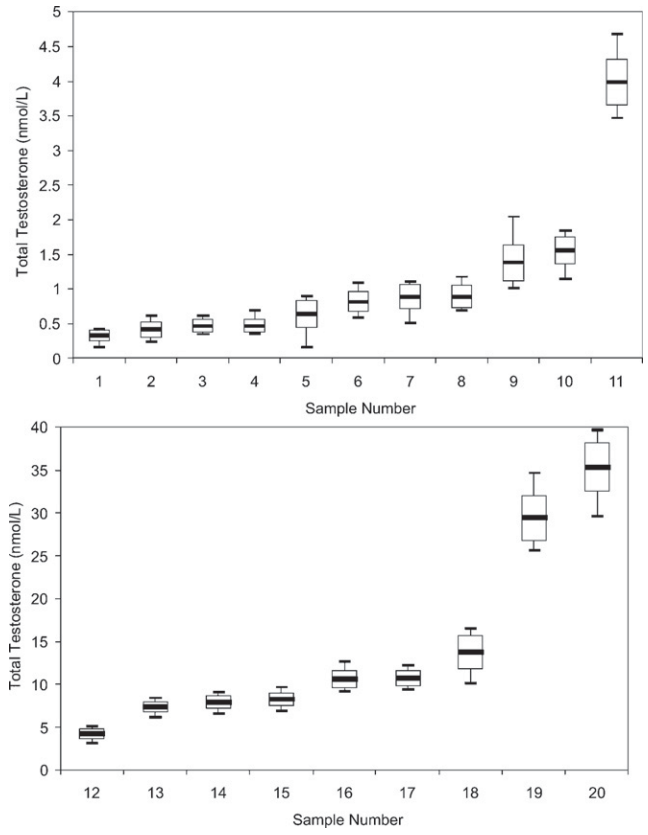

Fig. 1. Mean, standard deviation, minimum and maximum testosterone concentrations for each sample reported by MS assays (upper panel: samples ≤3.47 nmol/L (100 ng/dL) and lower panel samples >3.47 nmol/L).

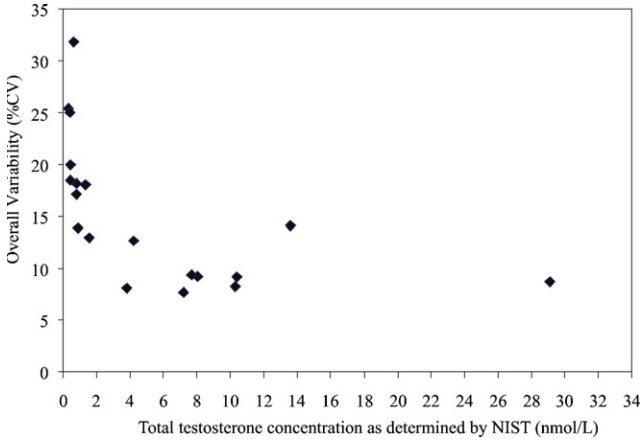

Fig. 2. Variability in individual testosterone results on individual samples performed with mass spectrometry assays.

The testosterone measurement results reported for each sample by all MS assays are summarized in Fig. 1. They ranged between 0.17–2.98 nmol/L (5–86 ng/dL) and 5.55–39.67 nmol/L (160–1,144 ng/dL) for female and male samples, respectively. The average ratios between highest and lowest reported concentration for samples with TT > 3.47 nmol/L (100 ng/dL) and TT ≤ 3.47 nmol/L were 1.4 (range: 1.3–1.6) and 2.3 (range: 1.4–5.4), respectively. For assay 8, results from four samples were above the reportable range. Results from all participating laboratories were reported in four different units (pg/ml, nmol/L, ng/dL, and ng/g).

The overall coefficient of variation of the MS assays was <15%CV for concentrations ≥1.53 nmol/L (≥44 ng/dL) and increased up to 33%CV at lower concentrations (Fig. 2). The within-run variability at 10.30 nmol/L (297 ng/dL) and 0.29 nmol/L (8.47 ng/dL) ranged between 1.40–11.36%CV and 2.52–25.58%CV, respectively (Table 2). Assays 5 and 8 were most precise and had a within-run imprecision that appears independent of the analyte concentration. The between-assay variability was the major contributor to the overall variability. However, at concentrations less than 5.0 nmol/L (144 ng/dL), the within-assay variability was also an important contributor to the overall variability.

The absolute values of the percent differences between the assays and the reference method were in average 11.4% (range of absolute values: 2.1–19.2%) (Fig. 3). They were smaller for concentrations >3.47 nmol/L (100 ng/dL) (average 7.5%, range of absolute values: 1.70–16.8%) than for concentrations ≤3.47 nmol/L (average: 15.5%, range of absolute values: 2.0–25.3%). The dispersion of data as indicated by the 95% confidence interval is higher at concentrations ≤3.47 nmol/L (100 ng/dL) as compared to concentrations >3.47 nmol/L (Table 3).

Weighted Deming regression analysis gave slopes ranging between 0.903 and 1.138. The slopes were significantly different

Table 2
Mean within-run imprecision of serum testosterone measurements determined on two samples measured in five replicates per sample on 2 different days.

	Sample A 0.29 nmol/L (8.47 ng/dL) (%CV)	Sample B 10.30 nmol/L (296 ng/dL) (%CV)
Assay 2	13.13	3.38
Assay 3	17.75	9.62
Assay 4	25.58	11.36
Assay 5	2.52	1.74
Assay 6	17.65	2.55
Assay 7	BR	1.40
Assay 8	2.67	2.19

BR: below reportable range.

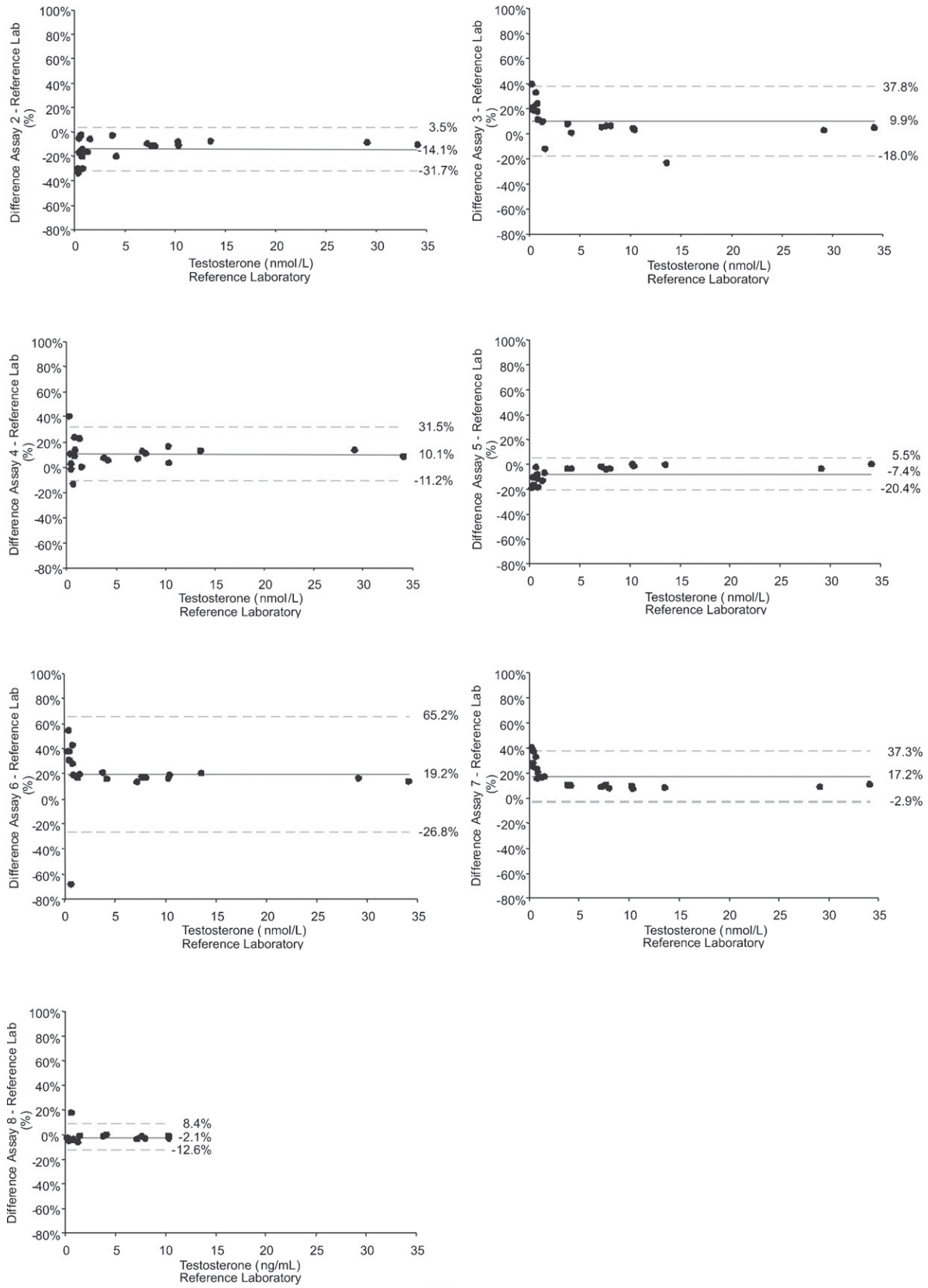

Fig. 3. Percent difference plot of mean results obtained on individual samples between individual assays and the reference method. Dashed lines: 95% limits of agreement.

from one in 6 assays, indicating proportional bias in these assays. The intercepts ranged between −0.053 and 0.107 nmol/L (−1.518 to 3.081 ng/dL) and were significantly different from zero in 4 assays, indicating a constant bias in these assays (Table 3). The correlation coefficients were >0.996 for all assays. The average recoveries calculated from samples mixed at different ratios ranged between 96.5 and 103% except for one assay that had an average recovery of 136%.

Table 3
Weighted Deming regression and bias plot analysis comparing laboratory results against those from the reference laboratory (assay 1).

	Weighted Deming regression analysis			Percent bias plot analysis		
	Intercept (nmol/L) (95%CI)	Slope (95%CI)	Correlation coefficient	Overall mean bias (95%CI)	Mean bias at TT > 3.47 nmol/L (95%CI)	Mean bias at TT ≤ 3.47 nmol/L (95%CI)
Assay 2	−0.048 (−0.090 to −0.007)	0.903 (0.871–0.935)	0.9997	−14.1 (−18.3 to −9.9)	−10.5 (−13.6 to −7.3)	−17.8 (−25.7 to −9.9)
Assay 3	0.107 (0.070–0.144)	1.001 (0.940–1.062)	0.9963	9.9 (3.3–16.6)	1.7 (−4.8 to 8.2)	18.2 (8.2–28.2)
Assay 4	0.021(−0.087 to 0.129)	1.082 (1.018–1.146)	0.9994	10.1 (5.0–15.3)	9.6 (6.6–12.6)	10.7 (−0.3 to 21.6)
Assay 5	−0.053 (−0.069 to −0.037)	0.974 (0.956–0.991)	0.9998	−7.4 (−10.5 to −4.3)	−2.3 (−3.4 to −1.1)	−12.6 (−16.6 to −8.7)
Assay 6	0.058 (−0.033 to 0.149)	1.138 (1.088–1.189)	0.9997	19.2 (8.2–30.2)	16.8 (15.0–18.6)	21.6 (−2.5 to 45.8)
Assay 7	0.093 (0.073–0.113)	1.086 (1.074–1.097)	0.9999	17.2 (12.4–22.0)	9.1 (8.3–9.9)	25.3 (19.1–31.5)
Assay 8	−0.001 (−0.015 to 0.013)	0.98 (0.961–0.998)	0.9998	−2.1 (−4.8 to 0.6)	−2.2 (−3.4 to −1.0)	−2.0 (−7.0 to 3.0)

4. Discussion

We assessed the variability of serum TT measurements obtained with routine MS assays using samples from both men and women. The methodologies used by the participating laboratories are the most commonly used sample preparation procedures and chromatography–MS technologies for TT measurements. The TT concentrations of the samples used in this study cover the normative range reported for TT in both men and women [35].

The lowest TT concentration in the sample provided (0.29 nmol/L [8.47 ng/dL] by the NIST assay) was close to the lower end of the calibration curves for some assays. Another assay was unable to quantitate elevated TT concentrations commonly observed in men. These findings show that MS assays cover different concentration ranges and therefore should not be assumed to be applicable for all concentration ranges needed for clinical and research applications in men, women, and children.

The variability of measurements reported from individual samples by all laboratories is surprisingly small for concentrations >3.47 nmol/L (100 ng/dL) considering the diversity of procedures, calibrators, and technologies used by the participating laboratories. This variability is substantially smaller than the variability reported among immunoassays [39]. The variability increases at lower concentrations commonly observed in women. This increase in variability does not improve greatly when excluding individual assays with high imprecision indicating that the overall variability at this concentration range is being contributed by all assays.

The inter-assay variability is the major contributor to the overall variability found in this study, especially at concentrations commonly observed in men. The within-assay variability only contributes a very small fraction to the overall variability at this concentration range indicating that different calibrations and sample preparations performed by the same laboratory are consistent. Additionally, assay imprecision becomes an important contributor to the overall variability at concentrations commonly observed in women.

The reasons for the increased imprecision at low TT concentrations are unknown. The presence of interfering substances in serum at low concentrations was reported on individual samples [32]. This occurrence may significantly affect precision and prevent reliable HPLC–MS/MS measurements. The effect of such interfering substances would result in different recoveries at different concentrations and in poor agreements among methods. However, the observed recoveries and between-method agreements do not indicate major problems with interfering substances. A more specific interference test would be required to confirm the presence of interfering compounds. Another study suggested that phospholipids can cause ion suppression and thus affect method performance [35]. The authors stated that this problem can be minimized by using liquid–liquid extraction with a highly non-polar solvent. The fact that the highest precision at both low and high concentrations is observed with assays using liquid–liquid extraction and with the GC–MS/MS assay where phospholipids do not interfere with the MS measurements seem to support this hypothesis. The laboratories reported results rounded to different decimals (from none up to three decimals). These differences can affect the calculated precisions, especially at low concentrations. Thus, the higher imprecision at low TT concentrations with some laboratories can be explained in part by differences in rounding of results.

The comparison of the individual assays to the NIST assay found significant mean differences of 10% or less for most methods, which is much smaller than those reported between immunoassays and MS assays [7,8,14]. Weighted Deming regression analysis found high correlations for all MS assays and small, but significant differences in slopes and intercepts for most assays. This could be explained, in part, by the use of different calibrators or calibrator preparations. Applying common calibrators may easily solve this problem. The mean differences between the investigated assays and the NIST assay are similar to those reported in other MS assay comparison studies [35,38]. However, the range of differences in mean percent and in slopes and intercepts of the weighted Deming regression is wider in this study than in a prior one [38]. Similarly, the within-run precisions of individual assays, especially at low concentrations are less consistent with the assays previously investigated compared with those reported in this study. These findings show that the performance of MS assays differs by laboratory and needs to be controlled and standardized.

This study included two samples with total testosterone values of 10.30 nmol/L (297 ng/dL) and 10.37 nmol/L (299 ng/dL) as determined by the reference assay, which are close to the cut point for androgen deficiency in men of 10.40 nmol/L (300 ng/dL) as stated by the Endocrine Society guideline [1]. The ranges of individual results reported on these samples were 5.55–12.21 nmol/L (160–352 ng/dL) and 6.86–12.62 nmol/L (198–364 ng/dL). The values from three assays (assay 4, 5, and 8) differed from those determined with the reference method by less than 5% and assay 6 by more than 14%, which is the total error limit based on biological variability [40]. These data indicate that some of the investigated MS assays can distinguish between normal and androgen deficient men even when values are close to the stated cut point.

According to Westgard [40], the allowable precision for TT assays should be 4.7%. Five assays (assay 2, and 5–8) would meet the precision criterion at 10.30 nmol/L (297 ng/dL) and two assays (assay 5, 8) at 0.29 nmol/L (8.47 ng/dL). The fact that some assays do meet this precision criterion at low as well as high TT concentrations indicates that the recommended precision can be met and assays not meeting this criterion can be improved. These criteria are based on the between- and within-subject variability and are calculated from data sources that used immunoassays to determine the biological variability. More data on biological variability obtained with new precise and accurate assays are needed to better define performance criteria based on biological variability.

The variety of units used to report testosterone results in this study indicate the need for harmonizing procedures to facilitate comparison of results across systems and to minimize errors during unit conversion.

In conclusion, the measurement results obtained by using the investigated MS assays show better comparability as compared to results obtained with different immunoassays especially at concentrations commonly observed in men [7,8,14,39]. Thus the applicability of the assays for specific clinical or research questions is not uniform across assays. Improvements in the performance of MS assays appear achievable through assay standardization activities such as provision of common calibrators, accuracy-based external quality assurance programs and other activities directed towards achieving a certain standard assay performance. Therefore, the standardization activities performed by CDC will address these issues and will include MS assays as well as immunoassays. Improvements of individual assays as a result of this study have already been made by some laboratories such as widening the measurement range to be able to measure samples with high TT content or using of confirmation ions to increase the assay specificity (personal communications).

Acknowledgements

Funding for this project is provided by Solvay Pharmaceutical through the CDC Foundation. The Division of Laboratory Sciences at the National Center for Environmental Health and the Division of Cancer Prevention and Control at the National Center for Chronic Disease Prevention and Health Promotion also contributed to this project. We would like to thank Dr. Sam Caudill for his assistance with the SAS calculations, and Dr. Julianne Bothelo, CDC and Christopher Shacklady, CDC for their contributions to manuscript preparation.

References

[1] Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2006;91:1995–2010.
[2] Scher HI, Halabi S, Tannock I, Morris M, Sternberg CN, Carducci MA, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol 2008;26:1159–84.
[3] Kanayama G, Amiaz R, Seidman S, Pope Jr HG. Testosterone supplementation for depressed men: current research and suggested treatment guidelines. Exp Clin Psychopharmacol 2007;15:529–38.
[4] Nieschlag E, Swerdloff R, Behre HM, Gooren LJ, Kaufman JM, Legros JJ, et al. Investigation, treatment and monitoring of late-onset hypogonadism in males. ISA, ISSAM, and EAU recommendations. Eur Urol 2005;48:1–4.
[5] American Association of Clinical Endocrinologists Medical Guidelines for clinical practice for the evaluation and treatment of hypogonadism in adult male patients—2002 update. Endocr Pract 2002;8:440–56.
[6] Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an androgen excess society guideline. J Clin Endocrinol Metab 2006;91:4237–45.
[7] Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid-chromatography–tandem mass spectrometry. J Clin Endocrinol Metab 2004;89:534–43.
[8] Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, et al. Testosterone measured by 10 immunoassays and by isotope dilution gas chromatography-mass spectrometry in sera from 116 men, women and children. Clin Chem 2003;49:1381–95.
[9] Van Uytfanghe K, Stöckl D, Kaufman JM, Fiers T, De Leenheer A, Thienpont LM. Validation of 5 routine assays for free testosterone with a candidate reference measurement procedure based on ultrafiltration and isotope dilution-gas chromatography–mass spectrometry. Clin Biochem 2005;38:253–61.
[10] Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for estimation of free testosterone in serum. J Clin Endocrinol Metab 1999;82:3666–72.
[11] Fritz KS, McKean AJS, Nelson JC, Wilcox RB. Analog-based free testosterone test results linked to total testosterone concentrations, not free testosterone concentrations. Clin Chem 2008;54:512–6.
[12] Moal V, Mathieu E, Reynier P, Malthiery Y, Gallois Y. Low serum testosterone assayed by liquid chromatography–tandem mass spectrometr. Comparison with five immunoassay techniques. Clin Chim Acta 2007;386:12–9.
[13] Hsing AW, Stanczyk FZ, Belanger A, Schroeder P, Chang L, Falk RT, et al. Reproducibility of serum sex steroid hormone assays in men by RIA and mass spectrometry. Cancer Epidemiol Biomarkers Prev 2007;16:1004–8.
[14] Kane J, Middle J, Cawood M. Measurement of serum testosterone in women; what should we do? Anal Clin Biochem 2007;44:5–15.
[15] Stanczyk FZ, Lee JS, Santen RJ. Standardization of steroid hormone assays: why, how, and when? Cancer Epidemiol Biomarkers Prev 2007;16:1713–9.
[16] Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? Clin Chem 2003;49:1250–1.
[17] Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an endocrine society position statement. J Clin Endocrinol Metab 2007;92:405–13.
[18] Rosner W. An extraordinarily inaccurate assay for free testosterone is still with us. J Clin Endocrinol Metab 2001;86:2903.
[19] Rosner W. Errors in the measurement of plasma free testosterone. J Clin Endocrinol Metab 1997;82:2014–5.
[20] Wierman ME, Basson R, Davis SR, Khosla S, Miller KK, Rosner W, et al. Androgen therapy in women: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2006;91:3697–710.
[21] Swerdloff RS, Wang C. Free testosterone measurement by the analog displacement direct assay: old concerns and new evidence. Clin Chem 2008;54:458–60.
[22] Stanczyk FZ. Measurement of androgens in women. Semin Reprod Med 2006;24:78–85.
[23] Matsumoto AM, Bremner WJ. Serum testosterone assays—accuracy matters. J Clin Endocrinol Metab 2004;98:520–4.
[24] Speiser PW. Interpretation of pediatric endocrine laboratory tests: pitfalls in steroid hormone measurements and genotyping. Pediatr Endocrinol Rev 2007;5S1:578–83.
[25] Albrecht L, Styne D. Laboratory testing of gonadal steroids in children. Pediatr Endocrinol Rev 2007;5S1:599–607.
[26] Thienpont LM, De Brabandere VI, Stockl D, De Leenheer AP. Use of cyclodextrins for prepurification of progesterone and testosterone from human serum prior to determination with isotope dilution gas chromatography/mass spectrometry. Anal Chem 1994;66:4116–9.
[27] Thienpont LM, Van Nieuwenhove B, Stockl D, Reinauer H, De Leenheer AP. Determination of reference method values by isotope dilution-gas chromatography/mass spectrometry: a five years' experience of two European Reference Laboratories. Eur J Clin Chem Clin Biochem 1996;34:853–60.
[28] Siekmann L. Determination of steroid hormones by the use of isotope dilution—mass spectrometry: a definitive method in clinical chemistry. J Steroid Biochem 1979;11:117–23.
[29] Wolthers BG, Kraan GP. Clinical applications of gas chromatography and gas chromatography–mass spectrometry of steroids. J Chromatogr A 1999;843:247–74.
[30] Tai SSC, Xu B, Welch MJ, Phinney KW. Development and evaluation of a candidate reference measurement procedure for the determination of testosterone in human serum using isotope dilution liquid chromatography/tandem mass spectrometry. Anal Bioanal Chem 2007;388:1087–94.
[31] Vogeser M, Parhofer KG. Liquid chromatography tandem-mass spectrometry (LC–MS/MS)—technique and applications in endocrinology. Exp Clin Endocrinol Diabetes 2007;115:559–70.
[32] Cawood ML, Field HP, Ford CG, Gillingwater S, Kicman A, Cowan D, et al. Testosterone measurement by isotope-dilution liquid chromatography–tandem mass spectrometry: validation of a method for routine clinical practice. Clin Chem 2005;51:1472–9.
[33] Rauh M, Groschl M, Rascher W, Dorr HG. Automated, fast and sensitive quantification of 17 alpha-hydroxy-progesterone, androstenedione and testosterone by tandem mass spectrometry with on-line extraction. Clin Chem Lab Med 2006;71:450–8.
[34] Vicente FB, Smith FA, Sierra R, Wang S. Measurement of serum testosterone using high-performance liquid chromatography/tandem mass spectrometry. Clin Chem Lab Med 2006;44:70–5.
[35] Kushnir MM, Rockwood AL, Roberts WL, Pattison EG, Bunker AM, Fitzgerald RL, et al. Performance characteristics of a novel tandem mass spectrometry assay for serum testosterone. Clin Chem 2006;52:120–8.
[36] Gallagher LM, Owen LJ, Keevil BG. Simultaneous determination of androstenedione and testosterone in human serum by liquid chromatography–tandem mass spectrometry. Ann Clin Biochem 2007;44:48–56.
[37] Turpeinen U, Linko S, Itkonen O, Hämäläinen E. Determination of testosterone in serum by liquid chromatography–tandem mass spectrometry. Scand J Clin Lab Invest 2008;68:50–7.
[38] Thienpont LM, Van Uytfanghe K, Blincko S, Ramsay CS, Xie H, Doss RC, Keevil BG, Owen LJ, Rockwood AL, Kushnir MM, Chun KY, Chandler DW, Field HP, Sluss PM. State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry. Clin Chem 2008;54:1290–7.
[39] Steinberger E, Ayala C, Hsi B, Smith KD, Rodriguez-Rigau LJ, Weidman ER, et al. Utilization of commercial laboratory results in management of hyperandrogenism in women. Endocr Pract 1998;4:1–10.
[40] Desirable specifications for total error, imprecision, and bias, derived from biologic variation [online] [cited April 16, 2008]. Available from URL: http://www.westgard.com/biodatabase1.htm.